

EXHIBIT A

[PROPOSED] ORDER RE: CONVERSATION DATA SAMPLING

WHEREAS, the Court has presented the “hypothetical[]” that “ChatGPT user[s]” who use ChatGPT to “get around [a] pay wall” might delete their conversations to avoid potential liability, 1/22/2025 Hr’g Tr. at 38:14–20, and that this behavior may result in a material difference between the conversation data OpenAI has retained and the conversation data that OpenAI does not retain on account of its commitment to honor user-initiated deletions, MDL ECF 42 at 5.

WHEREAS, the Court on May 16, 2025 requested that the parties propose a “means to determine” whether there is such a difference that might be relevant to this litigation. *Id.* at 5.

WHEREAS, the Court subsequently endorsed OpenAI’s¹ proposal to undertake “some form of sampling that enables us to compare whether there’s a . . . difference in the hit rate in these two different pools of data,” 5/27/2025 Afternoon Hr’g Tr. at 43:11–18, and confirmed that the Court was “absolutely not . . . interested in” ordering OpenAI to “produce[] to the plaintiffs” any data that OpenAI began retaining in compliance with the Court’s May 13, 2025 order, *id.* at 3:18–25; and

WHEREAS, on May 29, 2025, the Court ordered the parties to begin sampling the “30-day tables of consumer output log data” and ordered OpenAI to submit a proposed order regarding the sampling process, consistent with the discussion at the May 27 hearing, MDL ECF 79 ¶ 2;

ORDERED, that the following procedures shall apply to the sampling of the consumer output log data set forth in the Court’s May 29, 2025 Order:

1. As soon as practicable following entry of this Order, OpenAI shall generate data samples using a simple random sampling methodology as follows:

OpenAI’s Proposal	News Plaintiffs’ Proposal
a. Each sample shall consist of the contents of individual rows in the table of consumer output log data that OpenAI has	a. Each sample shall consist of the contents of individual rows in the table of consumer or API output log data that OpenAI has

¹ “OpenAI” shall refer to Defendants OpenAI, Inc., OpenAI OpCo, LLC, OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo, LLC, OpenAI Global LLC, OAI Corporation, and OpenAI Holdings, LLC. “News Plaintiffs” shall collectively refer to Plaintiffs The New York Times Company; Daily News, LP; Chicago Tribune Company, LLC; Orlando Sentinel Communications Company, LLC; Sun-Sentinel Company, LLC; San Jose Mercury-News, LLC; DP Media Network, LLC; ORB Publishing, LLC; Northwest Publications, LLC; and the Center for Investigative Reporting.

<p>segregated and preserved in response to the Court’s May 13, 2025 Order, MDL ECF No. 33 (the “Preservation Order”), where each row contains, <i>inter alia</i>, individual text prompts and outputs for requests generated by users whom OpenAI reasonably believes are located in the United States and corresponding text outputs generated by ChatGPT Free, Plus, or Pro using the models GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, or GPT-4o Mini, between April 14, 2025 and May 14, 2025 (hereinafter, “Conversation Data”)</p>	<p>segregated and preserved in response to the Court’s May 13, 2025 Order, MDL ECF No. 33 (the “Preservation Order”), or that OpenAI has otherwise retained, where each row contains all of the data fields ordinarily collected for consumer or API output log data, including, <i>inter alia</i>, individual text prompts for requests generated by users, corresponding text outputs generated by ChatGPT Free, Plus, or Pro or the API for all of the models at issue across all actions in MDL Case No. 25-md-3143 (SHS) (OTW), the intermediate queries and responses resulting from retrieval augmented generation (“RAG”) functionality, including any URLs and excerpts referenced in the RAG response, and metadata associated with each conversation such as classifiers (hereinafter, “Conversation Data”).</p>
---	---

- b. OpenAI shall use best efforts to draw independent and identically distributed samples from the sample populations described below:

OpenAI’s Proposal	News Plaintiffs’ Proposal
<p>i. The sample population for the first sample (“Control Sample”) shall be all Conversation Data that (i) was not generated through the “Temporary Chat” feature and (ii) is not subject to a user-initiated deletion. <i>See</i> MDL ECF 66 ¶ 3.</p> <p>ii. The sample population for the second sample (“Test Sample”) shall consist entirely of Conversation Data that was either (i) generated through the “Temporary Chat” feature or (ii) is</p>	<p>i. The sample population for the first sample shall be all ChatGPT data that is ordinarily retained from the time-period of April 14, 2025 through May 14, 2025, i.e., data that (i) was not generated through the “Temporary Chat” feature and (ii) is not subject to a user-initiated deletion..</p> <p>ii. The sample population for the second sample shall be all ChatGPT data subject to a user-initiated deletion that OpenAI has been ordered to retain from the time-period of April 14, 2025 through May 14, 2025.</p> <p>iii. The sample population for the third</p>

<p>subject to a user-initiated deletion. See MDL ECF 66 ¶ 3.</p>	<p>sample shall be all ChatGPT data generated through the “Temporary Chat” feature that OpenAI has been ordered to retain from the time-period of April 14, 2025 through May 14, 2025.</p> <p>iv. The sample population for the fourth sample shall be all ChatGPT data OpenAI has retained from April 14, 2023 through May 13, 2023, before any of the <i>MDL</i> lawsuits were initiated.</p> <p>v. The sample population for the fifth sample shall be all ChatGPT data OpenAI has retained from January 1, 2024, through January 31, 2024, the first month after The Times filed its Complaint.</p> <p>vi. The sample population for the sixth sample shall be all API data OpenAI has been ordered to retain from the time-period of April 14, 2025 through May 14, 2025.</p>
--	--

- c. Solely for purposes of complying with this Order, the size of each data sample shall be five (5) million rows of Conversation Data.
 - d. OpenAI shall undertake best efforts to anonymize the data in the Samples prior to any use of the data pursuant to this Protocol.
2. Within seven (7) days of generating the samples, OpenAI shall run keyword searches on each sample using the search terms listed on Appendix A across the data fields specified in Section 1(a).
 3. After running the searches in the foregoing paragraph, OpenAI will provide News Plaintiffs with

OpenAI’s Proposal	News Plaintiffs’ Proposal
. . . two tables (in a native spreadsheet format) containing hit counts for each search	. . . six tables (in a native spreadsheet format) containing hit counts for each

term. One table shall contain corresponding hit counts for the Control Sample and the other table shall contain corresponding hit counts for the Test Sample.	search term and news-related classifier or other metadata fields that indicate news-related use cases. Each table shall contain respective hit counts for Samples 1-6.
---	--

4. All data and analyses derived from the samples, including hit count tables, shall be designated as HIGHLY CONFIDENTIAL – OUTSIDE COUNSEL’S EYES ONLY pursuant to the operative Protective Order in this action.
5. All data and analyses derived from the samples, including hit count tables, may only be used for the sole purpose of determining the existence of any material difference between the Samples, as contemplated in the May 29, 2025 Order.
6. Seven (7) days after OpenAI provides the hit count tables for the Samples described above, OpenAI and News Plaintiffs shall submit a joint letter of no more than three (3) pages (no more than one-and-a-half (1.5) pages per side) that attaches the hit count tables and presents each side’s position regarding whether there exists any material difference in hit counts between the samples.
7. If News Plaintiffs’ findings submitted in paragraph 6 above do not purport to show the existence of a material difference between the Samples, the Preservation Order shall be vacated. If, however, OpenAI and News Plaintiffs submit disputed findings as to the existence of a material difference between the Samples, and that dispute is resolved by the Court in OpenAI’s favor, the Preservation Order shall thereafter be vacated,

OpenAI’s Proposal	News Plaintiffs’ Proposal
and OpenAI may recover any costs associated with its retention of the consumer output log data that OpenAI has segregated and preserved in response to the Preservation Order that it incurred from the date of the joint letter in paragraph 6 above.	N/A